

講演ノート

明確な患者利益の追求における 臨床疫学役割

◆ A note on lecture ◆

The Role of Clinical Epidemiology in the Search of Tangible Benefits

The goal in clinical decision-making is to obtain tangible patient benefits. The word tangible indicates that the end result of an intervention should be an outcome that patients can perceive. Such outcomes are referred to as clinically relevant outcomes or true endpoints. The clinically relevant outcomes in dentistry are saving teeth and improving dental related quality of life by improving esthetics, minimizing pain, or improving chewing function. In dentistry, there is almost no clinical research to assist clinicians in determining which clinical decisions are associated with tangible patients benefits.

Dental clinical research has focused mostly on the study of intangible outcomes; outcomes that patients cannot perceive. Intangible outcomes measures are referred to as surrogate endpoints or measures of disease process. Examples of surrogate endpoints are changes in probing pocket depths, changes in *Streptococcus mutans* levels, or changes in cephalographic anatomical landmarks. It has been a common and pervasive assumption in dental research that the study of factors that influence clinically relevant outcomes was unnecessary. In the USA and Europe, billions of dollars have been spent on the study of intangible outcomes, and yet almost none of these resources were focused on the study of clinically relevant outcomes.

Such a one-sided approach to clinical research can be dangerous; clinical interventions that influence intangible outcome measures do not necessary influence tangible outcome measures. Improvements in intangible outcome measures may result in real harm, or vice-versa. For instance, once widely used interventions that successfully lowered lipid levels, increased immune function, or normalized heart rhythms, have now been shown to increase mortality. In such instances, use of intangible outcome measures led to false-positive results; conclusions that interventions had positive effects, while in fact the interventions were harmful. Similarly, interventions have been identified that had no impact on 'important' intangible outcomes measures, and yet provided real benefits. In such instances, intangible outcome measures led to false-negative results; conclusions that interventions were ineffective, while in fact the interventions were beneficial.

If we want to stay true to our goal of providing tangible patient benefits, clinical epidemiology will have to start playing an increasingly important role in dental research. Clinical epidemiology is defined as the study of the factors that influence tangible patient outcomes. Clinical epidemiology can assist clinicians in providing reliable answers to important and controversial clinical questions related to etiology, diagnosis and treatment.

J Health Care Dent 2001; 3: 33-49.

フィリップ・フジョー

Philippe P. Hujoel, MSD, PhD, MS

ワシントン大学歯学部および公衆衛生学部準教授

Associate Professor, with joint appointments in the School of Dentistry and in the School of Public Health at the University of Washington

Warren G. Magnuson Health Sciences Center, Suite B-509, Box 357475, USA

訳読：内藤 徹 Toru NAITO
九州歯科大学

キーワード： true endpoint
surrogate endpoint
clinical epidemiology
tangible patient benefit

はじめに

ほとんどの場合、歯科の研究は、生物学的な疾患のメカニズムの理解について焦点を当てています。歯の喪失やQOLの低下といったシンプルで臨床に深いかかわりのある歯科疾患のエンドポイント^{注1)}についての研究はきわめて少ないのです。20世紀の医学の進歩をもたらした大多数の研究は、シンプルで臨床に深いかかわりのあるエンドポイントに基づく研究でした。生物学的なメカニズムは、むしろ治療法が確立した後何年も経ってから解明されたのです。

シンプルで臨床に深いかかわりのあるエンドポイントを用いた歯科疾患の研究は、多くの他の慢性疾患において示されたのと同じ期待を抱かせます。そのような歯科の研究の始まりは、歯科疾患の疾患コードの導入によって大いに助けられることでしょう。

有名な歯周病学者であり臨床家であったDr. Saul Shlugerは、1980年末、多くの優れた免疫学者の存在が歯周病の病因の信頼に足る証明をもたらす、治療法に帰結するに違いないという大きな希望をもっていました。彼自身、重度の歯周疾患に悩まされていたのです。今にして言えることですが、私たちは今日、彼の深刻な歯周疾患の原因の8割方は彼の喫煙癖にあったとすることができます(タバコを喫うDr. Shlugerの写真を示す)。喫煙習慣が歯周病の主たる原因であることの発見は、生物学的なメカニズムの研究によるものではありません。

シンプルで臨床に深いかかわりのあるエンドポイントに基づく臨床疫学的研究が、喫煙習慣が歯周疾患の大きな原因であることの強い根拠を提供したのです。

そこで今日は、疾患の原因を探する場合でも、治療法、診断法を検討する場合のいずれにおいても、はっきりと認知できる患者利益に焦点を当てて研究することが必要であるということ述べてみたいと思います。歯科疾患の研究は、主に生物学的、免疫学的、解剖学的な指標に焦点が当てられてきて、患者にとっての有用性には焦点が当てられてきませんでした。もし医学の分野でも同じようなアプローチがとられてきたとしたならば、私たちはまだ喫煙が肺ガンの原因であるとか、あるいは汚い水がコレラの原因であるとか、あるいは病気を治すと思われていた薬がかえって死亡率を高めてしまう、ということを知らないままだったでしょう。

1. 代用エンドポイントの罫

これからいくつかの医学の分野での例を挙げます。そして歯科の研究がどうなっているか、対比させて考えていきましょう。

図1は心拍曲線(心電図)です。まず、正常な心拍につづき心室性異所性収縮(期外収縮)が見られますが、心室に異常がありますと心室の期外収縮として表れます。無症候性に正常になるのですが、ここでは持続した頻脈となって、さらに心室細動に至り死亡しています。心室性異所性

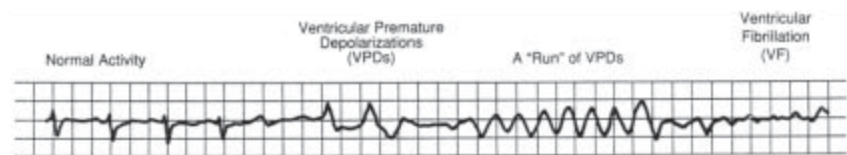


図1 心拍曲線(心電図)

注1) エンドポイントとは、治療結果や観察の結果を判定するのに用いる指標。たとえば歯周治療では、プロービング測定値や臨床的アタッチメントレベルがエンドポイントの一つであるが、この講義ではそれが真のエンドポイントではないことが繰り返し強調されている。

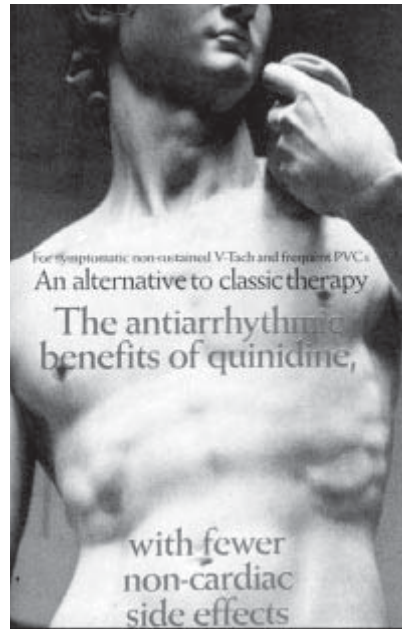


図2 『New England Medical Journal』に掲載されたキニジンの広告



図3 同, 抗不整脈作用の説明

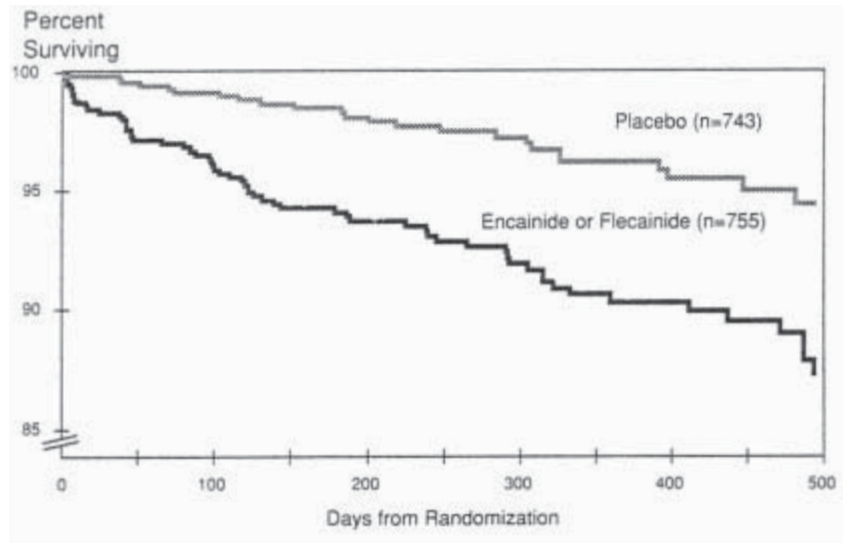


図4 臨床試験の結果

収縮を有する人は死亡率は3倍です。図2は『New England Medical Journal』(注:世界的に権威あるイギリスの医学雑誌)の1980年代後半に掲載されたキニジンの抗不整脈作用についての広告です。副作用がきわめて少ないと書いてあります。図3はENKAIDという抗不整脈薬の広告で、実際この薬を服用しますと、不整脈は正常に戻ります。この薬を服用すると心室性異所性収縮が70%減少し、また副作用による死亡例はわずか9%で

あると書かれています。一度この薬の服用を開始しますと、多くの人が継続的に服用をつづけます。つまりひどい副作用がないので、服用をつづけられるのです。しかし、不整脈が70%減少するということが、この治療が有効だということを十分に示す証拠と言えるのでしょうか。

確かにこの薬は不整脈を緩和することには効果があります。この薬を飲みますと心拍が正常に戻ります。また、疫学的な研究から、このよう

表1 臨床研究のエンドポイント

<p>代用エンドポイント</p> <p>— a surrogate endpoint is a laboratory sign or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint (Temple, 1995).</p> <p>真のエンドポイント</p> <p>— a true endpoint reflects unequivocal evidence of tangible patient benefit.</p>

な不整脈があると早期に死亡する確率が3倍になるということもわかっています。

たとえばこのような病態をもった友人がいたとします。あなたは、これだけの証拠をもって、この薬を推奨しますか。正常な心拍ということは一つの代用エンドポイント(surrogate endpoint)です。果して異所性収縮がなくなったということが、本当に真の効果と言えるのでしょうか。つまり心電図でいい結果が出るということは、延命につながるのでしょうか。

これを評価するために、無作為割り付け比較試験が行われました(N Engl J Med. 1991; 324)。この無作為割り付け比較試験では、1,489名を抗不整脈薬のグループとプラセボ(偽薬)群とにランダムに分けて被験者しました。この研究では、エンドポイントとして、心拍が正常であることではなく、延命効果があるということを見ました。

この臨床試験の結果を図4に示します。縦軸は生存率、横軸は投薬期間を示します。500日間この薬を服用したわけですが、500日後の生存率が示されています。私の12歳の娘に見せましたところ、すぐに理解できました。効果のないプラセボを投与された方がよかったわけです。「私もそっちの方がいいわ」と子供が言っていました。

プラセボ群は95%ぐらいの生存率を維持していますが、不整脈を正常化するには有効であるはずのこのエンカイド(ENKAID)投与群の生存率は、なんと90%を割って87%ぐらいになってしまっています。代用エン

ドポイントにおいて効果があるということと、それが真のエンドポイントにおいて効果があるということには大きな違いがあるのだということを申し上げたいのです。これが今日の私のお話しの主題です。

2. 真のエンドポイントと代用エンドポイント

臨床研究のエンドポイントには、真のエンドポイントと代用エンドポイントがあります(表1)。代用エンドポイントとは、患者が感じ、機能し、生存するというような直接認知することのできる患者にとって臨床的に意味のあるエンドポイントの代用として、臨床検査の値や身体的な兆候を用いるものです。代用エンドポイントにおいて生じる変化は、臨床的に意味のあるエンドポイントの反映であると受け取られます。真のエンドポイントとは、はっきりと分かる患者利益(tangible patient benefit)を決定的に示す証拠です。ここで“tangible patient benefit”ということばの意味を確認しておきましょう。ウエブスターによると“tangible”とは理性によって確実に同一視または実感できること、“benefit”とは幸福を増進するなにかで、“Outcomes of tangible patient benefit”は臨床的な実際的なアウトカム、真のエンドポイント、臨床的に意味深いエンドポイントと同義です。

歯科で、患者さんがはっきりと認知できるエンドポイントは何かといえますと二つしかないでしょう。歯の喪失と生活の質(QOL)です。代用エンドポイントにはどんなものがある

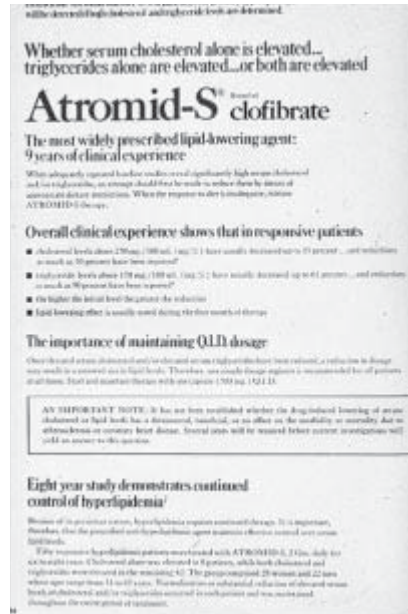


図5 『アメリカ医師会雑誌』に掲載された高脂血症に非常に有効とうたう薬の広告

るでしょう。研究のデータとして使われているもののほとんどです。たとえば、プロービングアタッチメントレベルです。

それでは代用エンドポイントと真のエンドポイントの違いをもう一つの例で示しましょう。

2-1 高脂血症薬 Clofibrate の例

図5は『アメリカ医師会雑誌』に掲載された広告で、高脂血症に非常に有効な薬の広告です。高脂血症薬としてアメリカで一番よく処方されている薬の一つです。コレステロールを下げ、脂質レベルを下げるのに非常に有効です。8年間この薬を使用し続けても効果は失われず、脂質は低いレベルに抑えられると書いてあります。図6の上のグラフは「コレステロールがこんなに下がる」ということを示しています。下のグラフは中性脂肪(triglycerides: 中性脂肪主成分)がどれだけ下がるかを示しています。

下の写真は、高脂血症では黄色腫という脂肪腫ができますが、この薬剤を飲むとこれが消えるという効果が示されています。このように代用エンドポイントであるコレステロール値や中性脂肪を下げるうえで非常

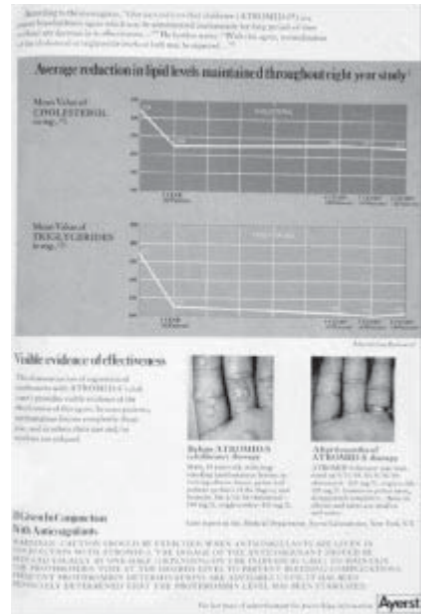


図6 同、薬の効果の解説

に有効であることが示されています。またコレステロールや中性脂肪が死亡率を上げるということも確かに示されています。

しかし、本当に知りたいのは、この薬を服用することによって長生きできるのかということです。たとえばこういう証拠を目の前につきつけられ、自分の脂質レベルやコレステロール値が高かったとします。皆さんはこの高脂血症薬を飲みたいと思いますか。

クロフィブレイト(Clofibrate)という高脂血症薬に関してWHOの呼びかけで共同研究が行われました(WHO: Lancet, 1984; 600)。Clofibrate群とプラセボ群に無作為に割り付けました。そして結果はこのClofibrate投与群の方が、何と47%も死亡率が高かったのです。このように代用エンドポイントだけを見ていると、効果について誤解してしまうことになります。

2-2 歯科の例—エイズ患者におけるアフタ性潰瘍

Jacobsonが1997年に『New England Medical Journal』に発表したエイズ患者におけるアフタ性潰瘍に対する

表2 エイズ患者におけるアフタ性潰瘍に対する Thalidomide の効果

TNF- α (pg/ml)	Sample size	Median change at 4 weeks	Inter quartile range
Placebo	22	0.9 pg/ml decrease	- 9.49 to 6.66
Thalidomide	20	4.4 pg/ml increase	0.14 to 8.83

表3 患者が認知できるエンドポイントによる評価

Response	Thalidomide	Placebo
Complete healing	55 % (16/29)	7 % (2/28)
Complete or partial healing	90 % (26/29)	25 % (7/28)

In addition, improved quality-of-life and weight gain associated with thalidomide.



図7 二つの処置方法を比較

左は有床義歯によって修復されたもの。右はインプラントを2本支台としてオーバーデンチャーにしたもの。

表4 QOLを指標にインプラントとフルデンチャーを比較

	Implant	Conventional
Functional limitations	14.4	18.5
Physical pain	12.5	18.7
Psychological comfort	7.1	10.3
Physical disability	12.1	15.8
Psychological disability	7.9	10.6
Social disability	5.5	7.0
Handicap	6.7	8.3

Thalidomide(サリドマイド)の効果に関する研究を見ましょう。アフタ性潰瘍はひどい疼痛を伴います。腫瘍壊死因子(TNF- α)がアフタ性潰瘍の原因であるとする研究から新薬がつけられました。そして新薬の有効性を調べるのに、TNF- α の産生を調べました。TNF- α という指標を使い、プラセボとサリドマイドを比較しました。ところがサリドマイドを服用しますと、TNF- α は増えてしまいます(表2)。サリドマイドに関して臨床試験をしている担当医師であればエイズ患者のアフタ性潰瘍ではサリドマイドは効かない、むしろ有害だということになってしまいます。

この薬剤は、有効なのでしょうか、無効なのでしょうか。そこで実際に患者が認知できるエンドポイントで評価してみました(表3)。これを見るとサリドマイド群では完治した率が55%、プラセボ群では7%しかありませんでした。完治したものと部分的に治癒したものを足すとサリドマイド群では90%になります。それに対してプラセボ群では25%にしかありませんでした。

この例は、代用エンドポイントで偽陰性の結果が出てしまう例です。代用エンドポイントだけを見ていたのでは、このサリドマイドという薬は役にたらず、害にすらなるというような結果になってしまいます。ところが実際に治癒するかどうか、患者の利益になるかどうかを見ると、非常に有効な薬であるということになったわけです。

2-3 有床義歯とインプラント支持の下顎義歯

もう一つ歯科の例を見てみましょう(図7)。写真の左は有床義歯による修復、右はインプラントを2本支台としてオーバーデンチャーにしたものです。たとえば皆さんが、二つの処置方法を比較するとき、どのようなエンドポイントを使いますか。代用エンドポイントですか、それとも真のエンドポイントを使った方がいいのでしょうか。

どのようなエンドポイントを使えば、実際に治療を受ける患者さんにとって利益になる結果が得られるのでしょうか。たとえば代用エンドポ

表5 医学、科学の研究に携わる人にとって非常に重要なメッセージ

『サイエンス』誌がコンタクトした臨床試験を行った人たちの大多数が、異口同音に述べていることが一つあります。それは、たとえば患者の生存のような明確に定義された臨床エンドポイントの代わりに、不適切な「代用」マーカーを用いたために、それがおそらくは臨床試験における最大の欠点となっているということです。

Science. 1994

臨床試験における代用エンドポイント：私たちは惑わされているのか？

Ann Intern Med. 1996; 605-613

私たちはこれまで臨床試験のアウトカムについての評価を行ってきました。そのアウトカムというものは、臨床家たちが、判断の拠り所として、抵当権から何からあらゆる財産を賭けてきたのです。しかし、私たちはそうした臨床家の行動には無作法にもいくばくかの驚きを感じていました。

David DeMets: Statistics in Medicine. 1996, 15: 1282

私たちはガンの化学療法についての臨床試験をこの20年間実施してきましたが、臨床試験の報告書にはごく最近まで、いわゆる症候、機能的な能力、QOL、もしくはヒトとラットを区別するような事柄について、何ら記載が見受けられませんでした。これは、私たちの時代の名誉にならないことです。

Alvan Feinstein: Statistics in Medicine. 1996, 15: 1282

イントとしては、ある一定の試験的な食物(人参など)を与えてそれを両方の群で咀嚼してもらう、十分に噛み砕くまでにどのくらい時間がかかるかを計るというのも一つの方法です。それに対して、患者さんが本当に自分で認識できるような利益があるかどうか、たとえば生活の質(QOL)が改善したかどうかを見た方がいいのでしょうか。

QOLを指標に比較したのが表4です。機能的な制限、たとえば咀嚼において制約があるか、従来のデンチャーよりもインプラントの方がいい結果が出ています。

このように代用エンドポイントと真のエンドポイントでは、随分結果が違ってくるといえる例はたくさんあります。

2-4 代用エンドポイントの導く結果

代用エンドポイントの導く結果について、Fleming TRのおもしろい文献^{注2)}があります。これは過去10年間に、どのような試験が行われて誤った結果を出したかということをもとめたものです。

ここに紹介されている例からザッと挙げていきましょう。まずAZTというエイズの治療薬がありますが、

CD-4の陽性細胞数はこの薬で上がりますが、死亡率は変わりません。骨粗鬆症に関して、フッ素は骨密度を上げますが、真のエンドポイントである骨折については、かえって骨が折れやすくなるという結果が出ています。ガンの治療薬、抗ガン剤でもいろいろな薬が出ていますが、抗ガン剤を服用すると腫瘍のサイズは小さくなるけれども、逆に死亡率は上がってしまいます。

第一の結論

歯科疾患では、この30～40年間は代用エンドポイントに焦点が当てられてきました。たとえばポケットの測定値に焦点が当てられるのに対して、歯の喪失やQOLなどは評価の対象とはなっていません。

代用エンドポイントを使った場合には、偽陽性または偽陰性の結果が出ますが、それをそのまま信じてしまうことがあります。偽陽性とは、代用エンドポイントで治療効果があるにもかかわらず実際は効果がないというケースです。具体的な例としてはエイズ、ガン、循環器疾患、骨粗鬆症などがあります。

これに対して偽陰性とは、代用エンドポイントでは効果がない、ある

注2) Fleming TR, DeMets DL: Surrogate end points in clinical trial: are we being misled?. Ann Intern Med, 1996;125:605-613.

表6 代用エンドポイントが患者利益になる
治療のエビデンスを提供することができない理由

- ・ Biological pathways are too complex
- ・ We commonly assume we know the causal pathway, while in reality our understanding is incomplete or wrong
- ・ Treatments influence outcomes through pathways we do not know
- ・ Don't rely on hypothesized causal pathways

いはむしろ害になるという結果が出ているにもかかわらず、真のエンドポイントでは効果があるという場合です。

医学、科学の研究に携わる人にとって非常に重要なメッセージを引用し(表5)、本日の一つ目の結論をしたいと思います。

歯科の治療において信頼できるエビデンスを得るためには、患者がはっきりと認知できる利益を指標として用いる研究を行わなくてはならないということです。このことを歯周病の研究で考えてみますと、これまで歯周治療に関して患者がはっきりと認知できるエンドポイントを用いた臨床試験は行われておりません。ここで言うはっきりと認知できるエンドポイントとは、歯の喪失であるとか、無歯顎になることであるとか、あるいは生活の質といったものです。これまでの歯周治療に関する臨床試験ではすべて代用エンドポイントが用いられています。このことを他の医学分野の研究と比較してみると、われわれは果たして歯周治療について、どの程度の信頼性のあるエビデンスをもっているのかという問いに向き合わざるをえません。

3. 代用エンドポイントが誤った結論を導くのはなぜか？

では代用エンドポイントが誤った結果を導き出してしまうのは、なぜなのでしょう。患者がはっきりと認知できる利益に比べて、代用エンドポイントが真実を私たちに教えてくれないのは、なぜなのでしょう。臨床研究において、代用エンドポイントはどのような役割を果たすべきで

しょうか。

代用エンドポイントが、なぜ患者利益になる治療のエビデンスを提供することができないのか、その理由を挙げてみました(表6)。まず第一に、慢性疾患の生物学的な原因結果の経路は、あまりにも複雑です。多くの慢性疾患について、われわれはその原因がわかっている、と考えているわけですが、現実にはわれわれの理解が間違っていたり不完全な場合が少なくありません。

私は四つの例を挙げましたが、患者が認知できるエンドポイントで評価を下すことができたと言っても、それがどういう機序によるものか、その経路についてはわれわれはまだ理解していません。たとえば最初に示した不整脈の例では、抗不整脈薬を使ってなぜ死亡率が上がったのかということは、現時点ではまだ分かっていません。他の代用エンドポイントの研究でも、同じような例がたくさんあります。代用エンドポイントを見ると有効なように思われるのに、なぜ最終的なところで有害な結果が出てしまうのか、現時点でもまだ分かっていないことがたくさんあります。

ですから、原因結果の仮説に基づいた研究では、その治療が疑う余地なく患者の利益になるのだということは証明できません。

歯周外科の評価に関する研究の例を図8に示しますが、たとえばプロービング・アタッチメントレベルをエンドポイントとして評価するときには、アタッチメントレベルに変化を起こさずに歯の喪失に至るような他の経路はないと仮定しているわけです。たとえば根面カリエスでも歯

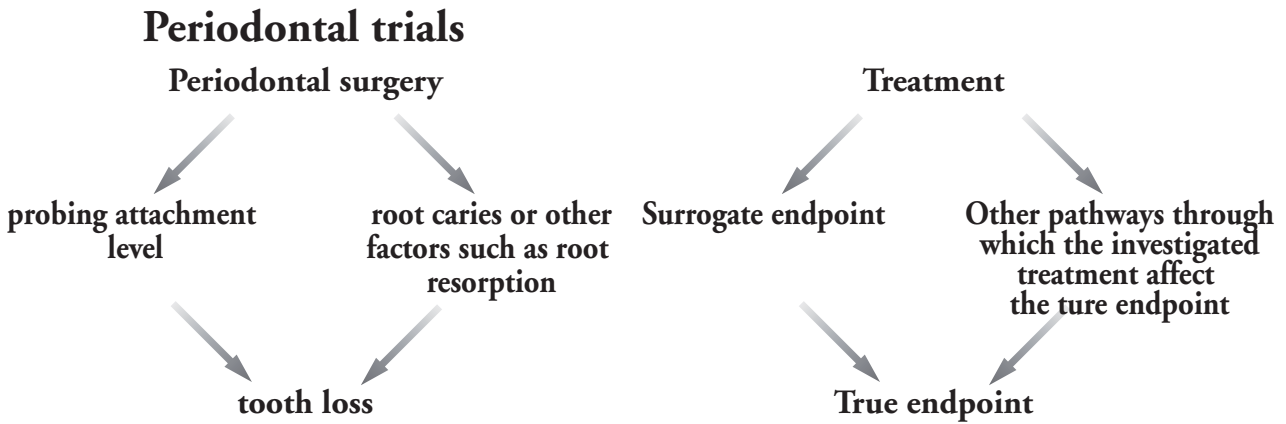


図8 歯周外科の評価に関する研究の例

の喪失が起こるわけですが、歯周外科とそのような他の要因とは全く関係がないのだと仮定しているわけです。ここで重要なことは、われわれがまだ気づいていないような多くの因子があるかもしれないということです。そうであれば、歯周外科が別の因子に影響を与えて歯の喪失につながってしまうかもしれないのです。

われわれが代用エンドポイントを用いてある治療の臨床評価をするときには、その治療によって代用エンドポイントに何らかの影響が出て、それによって真のエンドポイントの評価の代わりにすることができるのだということを仮定しているわけです。他にもいろいろな因子があって、ひょっとしたら治療によってそちらの因子に変化が出て、最終的な結果に影響するのかもしれないという可能性を見ていないわけです。

4. 新しい治療の評価に際しての代用エンドポイントの役割

では、代用エンドポイントの役割はどこにあるのでしょうか。それは、毎年のように開発されている新しい治療方法の探索です。たとえば歯周治療の分野では、皆さんよくご存知のように毎年のように新しい治療方法が発表されます(表7)。それらの一つ一つについて、患者がはっきりと認知できるような利益を評価する

だけのリソースも資金もありません。認知できる患者利益をエンドポイントとして臨床試験を行うと非常に高価になります。

たとえば心疾患の例ですが、ある抗高血圧薬がどの程度血圧を下げるかを評価をする場合、この血圧というのは代用エンドポイントです。このような研究であれば、臨床試験の期間は1~2年、対象患者数は200人くらい、臨床試験のコストはそんなに高くありません(表8)。これに対して同じ薬で患者がはっきりと認知できる利益、たとえば心臓発作の発生率で評価しようとする、臨床試験の期間は5年間と長くなり、対象患者数も25,000人となって、非常にお金のかかる臨床試験になります。

歯科の例で見てみましょう(表9)。たとえばクロルヘキシジンがストレプトコッカス・ミュータンスの菌数を30%減らすということを証明するための臨床試験であれば、期間は4週間、患者数は35人となります。これに対してははっきりと認知できる患者利益、たとえば患者の歯の喪失が減ることを証明しようと思えば、臨床試験の期間としては5年間、対象患者は約1,100人が必要です。

言うまでもなく新しく出てくる薬、治療法一つひとつについて、すべて認知できる患者利益を評価する臨床試験を行うということは不可能です。ですから代用エンドポイントを使った臨床試験は、いわばスクリーニン

表7 歯周組織再生療法に関して発表された新しい治療法

<p>barrier: Goretex, Temporary synthetic skin, Bio-resorbable type I collagen membrane (Perio-Barrier), Microfibrillar collagen hemostat (Avitene), FD dura mater allograft, Irradiated cross-linked human collagen membrane, collagen membrane, Autogenous Periosteal grafts.</p> <p>filler: β-TCP/Plaster/Dox, β-TCP+HA:(BCP), sintered HA (Durapatite), Granular porous HA (Interpore 200), FD Decalcified allogeneic bone (DFDBA), FD allo bo (FDBA), FD allo bo + Tetracycline, allogenic demineralized dentin, Autogenous bone particles, Autolysed antigen-extracted</p>	<p>allogeneic bone, Collagen gel, HTR polymer, Irradiated human collagen, b-TCP (Synthograft)/TTC, β-TCP /collagen gel (5:1), Granular porous HA (Interpore 200), Periograf (HA), Granular porous HA (Interpore 200), AAA bone+collagen gel, Osteogenin+ DFBA, Osteogenin+collaplug, collaplug, Lyophilized allogenic dura mater (Iyodura), cartilage allografts, Periograf (HA)/TTC, FDBA(4):TET(1), Block porous HA (Interpore 200), Human allografts iliac bone+marrow, Orthomatrix HA-500, Cartilage Allograft, Lyophilized bone allografts.</p>
---	---

表8 エンドポイントの違いによる臨床試験のコスト (抗高血圧薬)

- Does a novel medication result in a **blood pressure** reduction?
 - study duration: 1-2 years
 - sample size: 200
 - Cost: moderate
- Does a novel medication reduce **the incidence of stroke**?
 - study duration: 5 years
 - sample size: 25,000 patients
 - Cost: expensive

表9 エンドポイントの違いによる臨床試験のコスト (クロルヘキシジン)

- Does CHX result in a 30% reduction in ***S. mutans***?
 - study duration: 4 weeks
 - sample size: 35 patients
 - Cost: moderate
- Does CHX result in a 30% reduction in **tooth loss** rate?
 - study duration: 5 years
 - sample size: 1,100 patients
 - Cost: expensive

表10 2種類の歯周治療を比較するエンドポイントの違いによる臨床試験の必要な期間とサンプル数

- To reliably detect that 2 periodontal treatments differ on probing attachment level measurements requires a few 100 subjects followed for 1 or 2 years.
- To reliably detect that 2 periodontal treatments differ by about 35% in their respective tooth loss rates, about 2,000 patients need to be followed for 5 years.

グとしてどの治療、どの薬が特に見込みがありそうかということをお互に分ける働きをします。代用エンドポイントを使った臨床試験であれば、期間が短くなり、対象患者数も少なくなり、コストも安くなります。これに対して真のエンドポイントを評価すると、期間も長く、対象患者数も多く、コストも非常に高くなります。ですからお金を出す側から考えると、真のエンドポイントの評価する臨床試験にはかなり抵抗があるというのも理解できます。

歯周炎の例で見て考えてみましょう(表10)。たとえば2種類の歯周治療を比較してプロービング・アタッチメントレベルに差があることを示すには100人ぐらいの患者を対象にして、1~2年フォローアップすればいいわけです。これに対して歯の喪

失率が35%違うということを証明するためには2,000人の患者を対象に、5年間フォローアップする必要があります。

第二の結論

臨床試験における代用エンドポイントの役割は、治療が影響を与える生物学的な経路について、あまりコストをかけずになおかつ短期間に探索することにあります。たとえばガンの新薬の臨床試験の場合では、患者に認知できる利益を評価の指標にした臨床試験が1%くらいありますが、私はこれと同じようなことが歯周治療の分野でも生ずることを願っています。すなわち代用エンドポイントを使った臨床試験をたくさん行い、そのなかの少数のものについては真のエンドポイントの評価する臨



図9 コレラの死者をプロットした地図(1854年5月)

表11 飲料水の会社の違いによるコレラの死亡者

Water supply	Number of houses	Deaths from cholera	Death rate (/ 10,000 houses)
Company A	40,046	1,263	315
Company B	256,423	98	37

床試験を行うという方法です。

つまりこれまで行われてきた、いわば探索的な歯周治療の臨床研究をガイドとして、患者利益につながる見込みがもっとも高い治療に的を絞った真のエンドポイントを評価する臨床試験を行うべきだと思います。

そこで本日の第二番目の結論になりますが、代用エンドポイントは真の患者利益にかかわる影響について、その根拠を高い信頼性をもって提供することはできません。代用エンドポイントの役割は、さまざまな新しい治療の可能性をスクリーニングすることです。

5. 病因論に関する研究

5-1 汚れた水とコレラ

今までのお話は、「治療」を評価するものでした。次に、病因論すなわち疾患の原因を理解することをテーマにしましょう。患者のはっきりと認知できる利益はどのような役割を果たすでしょうか。20世紀には、いろいろな臨床研究が行われ、医学の進歩をもたらしましたが、ランダムマークとなった研究は、いずれも患者が直接はっきりと認知できる利益に焦点をあてたものでした。どの例をとっても同じ特徴があります。

まず疫学研究者ならば、絶対に避

けては通れない研究があります。それはコレラと水質との関係を調べた研究です。この研究は、とても基本的な方法で行われました。図9の地図に黒くマークされている部分は、コレラによる死亡者を示しています。これを見ると非常に面白いことがわかります。これらの黒いマークは、ほとんど水源となる井戸の周囲にあります。そしてビールの醸造工場があります。その工場には150人の従業員がいたわけで、多くの死亡者が出てもよかったのに死亡者はゼロでした。ビールで救われたのです。このコレラによる死亡の原因を汚れた水に求めた研究は、次のような言葉で評価されています。

英国において150年間コレラの流行がない。これはジョン・スノウの鋭い観察と論理的な思考の結果であり、私たちはコレラから解放されている

これらが歯科と何の関係があるのか、歯周疾患と何の関係があるのかと思われるかもしれませんが、歯周病の病因論の研究と、このコレラの研究を比較してみますと、コレラのこの研究では、患者がはっきりと認知できるような利益を指標に観察をしたわけです。

表 12 肺ガンと喫煙との関係に関する研究

- ・ 59,600 questionnaires send out October 1951
- ・ 67 % (n = 40,564) replied
- ・ Deaths recorded through death certificates
- ・ March, 1954— a fist preliminary report was prepared

Smoking Dose	Death Rate due to Lung Cancer
Non-smokers	0.00
1-14 g	0.48
15-24 g	0.67
> 25 g	1.14

Doll and Hill: BMJ, 1954; 4877

表 13 血中の脂質のレベルと冠動脈性心疾患の関係

- ・ Framingham
 - 2,022 males
 - 1,689 females
- ・ Livermore
 - 1,961 males

Parameters	Incident Heart Disease
Cholesterol	Sign. Elevated
Systolic blood pressure	Sign. Elevated
Diastolic blood pressure	Sign. Elevated
Weight	Sign. Elevated

Gofman, et al: Circulation, 1966; 679

歯周炎に関して、どういう研究がなされてきたかというところ、コレラのこの研究と全く逆です。まず注目された指標は、患者が認知できないような項目でした。歯周疾患では、末期に歯の喪失がもたらされますが、それも無視してきました。コレラの場合には非常に簡単な指標を用いました。たとえば、微生物のメカニズムがどうだとか、なぜ水質が悪いと問題になるかという詳しいメカニズムではなく、水質だけを見たわけです。歯周炎においては、疾患の生物学的なメカニズムがわかっているということ的前提として非常に複雑な、たとえば歯肉炎指数あるいはその他の指数を指標としてきました。

このような疫学的に病因を探る研究においては、何かが原因になっているという仮説を立てないで、それを前提としないで研究した方がよいのです。そうすることによって非常にシンプルな関係、たとえば水質が悪いということとコレラとの関係が明らかになってきます。

表 11 は英国で飲料水を供給している A 社と B 社を比較してみると、A 社の水を使っている家では B 社の 9 倍もコレラによる死亡率が高いことが分かりました。この統計からコレラがこの水質と関係していることは明らかです。その水に何が含まれていて、どのようなメカニズムでコレラが発症しているかは、まったく分かっていません。しかし水とコレラ

との関係ははっきりしています。

5-2 喫煙と肺ガン

もう一つ有名な例を見てみましょう。表 12 は、肺ガンと喫煙との関係に関する研究です。この Doll and Hill の研究は、20 世紀の公衆衛生のランドマークとなった研究で、この研究により二人はサーの称号を受けています。この調査は、医師に対するアンケート調査の形で行われましたが、はっきりと認知できる患者の死亡で評価しました。さてその結果は、喫煙をしていないグループ (Non-smokers) は肺ガンによる死亡が 0 でした。それに対して、喫煙量を重量換算で集計していますが、喫煙量が増えるにつれて死亡率も高くなっていることがわかりました。

この肺ガンの研究でも、はっきりと認知できるエンドポイントを見えています。死んだか死んでいないか、YES か NO かという見方をしています。それに対して歯周炎の研究はどうだったでしょう。患者が認知できない代用エンドポイントを評価の指標としています。その疾患の末期の状態である歯の喪失は無視されています。

それから何に曝露されているのかということも、単純に喫煙ということに焦点をあてています。どういう生物学的なメカニズムで喫煙が肺ガンを起こすのか、あるいは全身疾患を起こすのかということは見えてい

表 14 B型肝炎と肝ガンの関係

<ul style="list-style-type: none"> ・ Most common malignant neoplasm in Asia ・ Suspected causes prior to 1981 <ul style="list-style-type: none"> — Aflatoxin 			
HBs status on recruitment	PHC or cirrhosis	Population at risk	PHC incidence
HBsAG-positive	57	3,454	1,158
HBsAG-negative	3	19,253	5

Beasley, et al.: Lancet, 1981; 1129

ないわけです。

ところが歯周病の研究では、全く反対のことが行われています。歯周病を起こす生物学的なメカニズムがわかっているかのような大前提のうえで研究が行われています。その一つがプラークです。まだプラークが歯周疾患を起こすのかどうか、結論が出ていないのに、それを前提として研究がデザインされています。しかも使われるエンドポイントは代用エンドポイントです。

5-3 血中の脂質レベルと冠動脈性心疾患

もう一つ例を挙げましょう。血中の脂質のレベルと冠動脈性心疾患の関係ですが、表 13 は最初のランドマークとなった1966年に発表された研究です。

FraminghamとLivermoreという町で約5,000人を対象にした研究です。この研究によってその後の一連の虚血性心疾患に関する研究が導かれたと言っても過言ではありません。ここでもはっきりと患者に認知できるあるいは死亡を評価の指標にして、コレステロール、あるいは収縮期血圧(最大血圧)、拡張期血圧(最小血圧)などが検討されています。

一つの分野のランドマークとなる研究というものは、何が疾患を引き起こす原因か、その証拠を見つけて、それを基に、そこから一連の研究が始まるのです。

ですから喫煙が肺ガンを起こす、汚い水がコレラを起こす、あるいは血中脂質の高いことが心疾患による

死亡を引き起こすというような、おおまかなことがわかって、そこから生物学的なメカニズムは何なのかという、より詳細な研究に入っていくべきであって、その逆であってはいけません。この脂質のレベルを見た研究でも、患者がはっきりと認知できる結果に焦点をあてています。それは末期の病態、死亡であり、心筋梗塞であったり、狭心症に焦点を当てています。

5-4 肝ガンとB型肝炎

最後の例を挙げましょう。20世紀における医学研究のランドマークとなる研究を見てきましたが、最後の例は、肝ガンとB型肝炎の例です。

1981年以前は、肝ガンの原因についてさまざまな説があり、議論がありました。

表 14 は Beasley らが1981年に行った研究の結果です。B型肝炎の抗原陽性の人は抗原が陰性の人に比べて肝ガンの発現率が250倍も高いということが示されました。いまでは、このウイルス感染によって肝炎になるということが肝ガンの唯一の原因であろうとされています。

ここでもやはり、患者にはっきりと認知できるエンドポイントを見えています。死亡です。

5-5 病因論の研究のまとめ

ここまで医学分野における主な四つの臨床研究を紹介しましたが、どの試験においてもサンプルサイズが非常に大きいということ。そしてどの研究においても患者がはっきりと認

表 15 歯周病の原因

Association	Oral hygiene and Periodontitis
Year	1963
Sample size	12 healthy subjects (~ 23 years old)
Endpoint	Gingival Index Plaque Index
Tangible Outcome	no

Löe, et al.: J Periodontol, 1965; 177

表 16 プラークと歯周病の関係

Dental assessment	Students	Tea Pickers
Plaque	1.5 (0.8)	2.5 (1.0)
Gingivitis	0.3 (0.3)	0.6 (0.4)
Periodontitis	3.9 %	4.2 %

Lancet, 1993; 342: 471-72

知できるエンドポイントを評価対象としています。

それでは疾患の原因を研究するに当たり、臨床的に深い意味のあるエンドポイントを評価することがどうして重要なのでしょうか。

まず20世紀の医学の歴史では、臨床に深い関連のあるエンドポイントを指標とすることによって疾患また死亡の原因が確認されました。栄養不良、汚れた水、喫煙などなどです。まずこれらの因果関係が最初に判明し、その生物学的なメカニズムの理解は、その後かなり後になってからわかっています。

たとえば水が汚いこととコレラでは、その両者の関係がわかってから30年経ってやっとビブリオコレラ菌が発見されました。喫煙と肺ガンに関しても、さきに示した最初の論文が出てからかなり経過して、1990年の後半になってやっとそのメカニズムがわかりつつあります。肝ガンに関しては、まだ現時点でも、なぜウイルス感染によって肝ガンが起こるのかそのメカニズムはわかっていません。疫学的な研究は、生物学的なメカニズムはわからなくても死亡の原因をつきとめることが可能なのです。歯周病の研究では、病因の生物学的なメカニズムについて仮説が立てられ、それにもとづいてより詳細なメカニズムを明らかにするためのたくさんの研究が行われています。はっきりと認知できるエンドポイントを使った疫学的な研究は、ほとんど試みられてきませんでした。ようやく最近になって喫煙と歯周疾患との関連が指摘されるようになるという事態になってしまったわけです。

われわれは歯周病の原因についてのわれわれの理解を過信して、疾患のプロセスにおける生物学的な指標の測定に焦点を当ててきたわけです。病因に焦点を当てた疫学的な研究、すなわち歯の喪失という最終的な結果に注目した研究はこれまでほとんど行われていません。

5-6 プラークと歯周病

表15は、歯周病の病因に関する研究で、非常に重要な役割を果たしている研究です。この研究が現在の歯周病研究のベースになったと言う人もいます。この研究は1965年に発表されたもので、12名の健康な被験者を対象に口腔内の清掃を行わなかったら歯肉炎指数(GI)がどう変化するかということ調べたものです。このような実験的な研究が、プラークと歯周炎に関係があるという仮説の基礎になったわけです。これに矛盾するようなエビデンスは無視されてきました。

たとえばこちら、表16は1993年に『Lancet』に発表された論文です。この研究では大学生と茶摘み労働者とを比較しています。まず、茶摘み労働者の方が大学生よりもプラークが多い。歯肉炎についても茶摘み労働者の方が多く、ところが歯周炎を評価すると差がありません。結論としてこの論文の著者はプラークコントロールを通じて歯周炎を予防しようとする試みは成果が得られないかもしれないと言っています。

私は、何もこの論文の著者と同じ意見だからこの論文を出してきたわけではありません。そうではなくて、他の医学分野の疾患で病因について

表 17 臨床研究の比較

Study Characteristics	Poor water (5-1)	Smoking (5-2)	Hep B (5-4)	Lipids (5-3)	plaque (5-6)
Tangible Endpoint	Yes	Yes	Yes	Yes	No
Biology Understood	No	No	No	No	-
Large Samples	Yes	Yes	Yes	Yes	No

言われているようなエビデンスが、歯周病の領域ではまだない、ということをお願いしたいのです。歯肉炎指数 (GI) も一つの代用エンドポイントですが、プラークであるとか、あるいは微生物学的評価、あるいは臨床的なさまざまな指標と歯周炎との関係はもうイヤというほど研究されてきました。これらの研究は「歯周病の原因はわかっている」「プラークがあるから歯周病になる」「プラークが多いから歯を喪失する」という認識に基づいて行われてきたものです。しかしその前提となる仮説を立証した研究はないのです。そのかわりに、なされるべき研究がなされず、たとえば喫煙が歯周炎における非常に重要なリスク因子であるということが、ごく最近(90年代)まで何十年にもわたって見逃されてきたということは非常に残念なことです。われわれが今後、患者の利益を評価するような臨床試験を行っていくことによって、歯周病の原因は明確になってくることでしょう。

第三の結論

これまで見てきたさまざまな臨床研究(表 17)を振り返ってみましょう。まず汚れた水の研究では、はっきりと認知できる明確なエンドポイントを使っていました。そして病因の生物学的な機序についてはわかっていたので、大きなサンプル数を使っています。他の三つの臨床研究、肺ガン、B 型肝炎、そして血中脂質についても同様です。これに対してプラークと歯周病の関係では、われわれははっきりと認知できるエンドポイントを評価項目としてこなかった、そしてサンプル数も大きくなかった、それでも生物学的なメカニ

ムについては、理解していると考えています。歯科の研究分野では代用エンドポイントを使った研究に何十億ドルもの資金が投資されてきました。アメリカでは毎月のように、新しい微生物が歯周病の原因菌として取り沙汰されています。そのほとんどすべての研究は、生物学的なメカニズムの仮説に基づいて行われています。そして患者が認知できる歯の喪失のような結果に焦点を当てた疫学的研究はほとんど行われておりません。

そこで今日の第三番目の結論になります。歯科疾患の病因について信頼性の高いエビデンスを得るためには、はっきりと認知できる患者利益を評価項目として用いる研究を行わなくてはならない。

6 疾患の診査

6-1 胸部レントゲン写真

では最後のトピックス、疾患の診断について話したいと思います。ここで重要な問いは、「その診査が患者にとっての明確な利益になりうるか」ということです。たとえば胸部レントゲン写真を撮ることで肺ガンの死亡率は減るのでしょうか。ここでも重要な問いとなるのは、胸部レントゲン写真による診査が患者にとっての明確な利益になるかということ です。たとえば、胸部レントゲン写真の診査で、肺ガンの死亡率は減るのでしょうか。胸部レントゲン写真を撮ったことで、肺ガンを検出できるかどうか、それだけでは意味のある問いにはなりません。患者として知りたいのは、胸部レントゲン診査を受けることによって「長生きにつながるのか、あるいは生活の質が高まる

のか」ということです。

この最初の問いの答えは、ノーです。信頼できる研究によると、定期的に胸部レントゲン検査を行うと、死亡率はむしろ高まります。

6-2 マンモグラフィ

次に同じような問いですが、マンモグラフィ（乳房のレントゲン検査）を行うことで、女性の死亡率は減るでしょうか。マンモグラフィによって、乳ガンが見つかるかどうかということは問題ではありません。そうではなくて、その検査を受ければ、「私のQOLは高まるのでしょうか。私は長生きできるのでしょうか」ということが、患者の知りたいことです。この2番目の問いの答えはイエスです。定期的にマンモグラフィの検査を受ければ、寿命は長くなります。

6-3 前立腺の表面抗原のスクリーニング試験

3番目の質問も同じようなことです。PSA、前立腺特異抗原のスクリーニング試験はどうでしょう。この試験によって前立腺ガンが見つかるかどうかではなくて、この試験を受けることによって「寿命が長くなるか、あるいはQOLがよくなるか」ということを知りたいわけです。

臨床研究の目的が、新しい治療方法を検討するときでも、あるいは疾患の原因を調べているものでも、あるいは診査の有効性について検討している場合でも、常に重要なのはそれが患者の生存あるいはQOLの改善に貢献するかどうかということです。

6-4 シミュレーション試験

歯周病の治療決定におけるIL-1の遺伝子診査の価値

歯科の多くの研究において、臨床的に深い意義のあるエンドポイント、たとえば歯の喪失でや患者の生活の質とかそういった事柄に、ある診査がどのような影響を与えるかということについては答えが出されていま

せん。

残念ながら現時点ではその診断が歯の喪失、あるいは患者のQOLにどのような影響を与えるかという研究は行われていないのですが、ワシントン州でいわば予備実験のようなことをやりましたので、その例をご紹介します。

ここで掲げる問いは、治療を始める前に歯周病の重症化を左右する遺伝的因子であるといわれるインターロイキン-1(IL-1)の遺伝子型を診査することに価値があるかという問題です。もちろん、もっとも確かな方法は、患者を二つのグループに無作為割り付けて、一方の群の人には遺伝子試験を受けてもらう、他方は受けてもらわない、結果的に歯の喪失がどうだったかという明確な評価指標で判定する方法です。ここではその理想的な試験ではなく、いわばシミュレーションを行いました。

IL-1の遺伝子型の検査は、重篤な歯周疾患になる可能性のある人を検出できるとされています。この試験が歯周治療の進行にどのような影響を与えるのか、またコストに関してはどのような影響を与えるのかを検討しました。

まずこの検査をします。IL-1の型を調べることにより、陽性か陰性かにまず分かれます。それからテストを行って偽陽性か偽陰性かを振り分け、最後の段階でこの検査結果から推奨される治療に従うか従わないかここでまた分かれます。そして歯周疾患が進行していく確率をシミュレーションしてみました。

まずこのIL-1陽性の人、つまり遺伝学的に重篤な歯周疾患になる素因をもっている人が30%と考えます。陽性者は、喫煙者と非喫煙者にわかれます。喫煙者だとわかれば、われわれはまず「タバコをやめなさい」と言うわけですが、その結果タバコを止める人と依然としてタバコを吸っている人に分かれます。そして最終的に、治療を受けるか受けないか、そのテストの結果から推奨される治

療に従うか、従わないかに分かれています。

シミュレーションモデルは、軽度あるいは中等度から始まって、そこから治療を受けて軽度あるいは中等度そのままメンテナンスに移行するかもしれないし、あるいは重度の状態になって、さらに重篤度がひどくなるかもしれません。それぞれのパターンの確率を考えてみます。

このようなシミュレーションを試みますと、最悪の場合、30万ドル近くコストが増えて、なおかつ重篤な歯周炎になる人が患者1,000人当たり3.6人増えるという結果が計算され、また、逆にコストが80万ドル以上減って、重篤な歯周炎の人も52.8%減少するという結果まで、非常にバラツキのある結論が出ました。ともかく、この検査が患者のはっきりと認知できる利益につながるかどうかについては、有用な結論を導き出すことができなかつたわけです。

なぜこのように不確実な結果が出たかと言いますと、第1にはテスト結果に患者がどの程度従うか、協力するか、その度合いが不明です。そして第2に非外科療法の効果がどの程度あるのか分からない。第3にIL-1の遺伝子型による疾患の進行の相対的リスクが明確ではないということです。

たとえば、ある種の人たちは、われわれがどんな治療をしようと全くその疾患の進行に変化がない、ということもあり得るわけです。

第四の結論

そして結論の第四番目になりますが、診査の価値について信頼性の高いエビデンスを得るためには、その診査が歯の喪失、あるいは患者のQOLにどのような影響を与えるかということを調べなくてはなりません。やはりはっきりと患者が認知できる利益を評価して、長期的にその患者にどのような利益がもたらされるのかということを検討しなければならぬのです。

おわりに

歯科疾患の疾患コードの作成が急務

今日の話の最初に引用したFeinstein Aの言葉を、少し変えてもう一度出したいと思います。

歯科の臨床試験は35年来行われてきたが、ごく最近まで、たとえば症状や、機能的な能力、QOLのようにヒトとラットを区別するような要因については、ほとんど報告がなかったということは、非常に残念なことであると言えます

21世紀に入って、認知できる患者利益を評価する臨床試験をスタートするのにいい時期になっています。これまで様々な慢性疾患の医学研究から学んだことを活用していけば、歯科の分野でも大きな進歩が期待できると思います。

治療であれ、あるいは疾患の原因の研究であれ、診査方法の評価であれ、臨床的に深い関連のあるエンドポイント、たとえば歯の喪失などの原因を明らかにし、あるいはそれを予防するような因子について研究をしなくてはなりません。そしてわれわれの中の多くの人々が、歯科の疾患コードの導入に向けて、仕事をしています。これも非常に重要なことです。

この疾患コードは、医学分野では17世紀からあります。この疾患コードがなかったら、たとえば喫煙が肺ガンを引き起こすとか、あるいは水の質が悪いとコレラになるとか、B型肝炎から肝ガンになるとか、そういった事実は明らかになっていなかったでしょう。そういった研究成果もこの疾患コードがあってこそ得られたわけです。ぜひとも歯科の疾患コードを急いで作らなくてはなりません。それがあって初めて治療を評価し、あるいは疾患の原因や診断のための試験などを評価することができるわけです。